# Introduction to Statistics I

**Instructor**: Jodin Morey     moreyj@lemoyne.edu

## Previous Lecture

- Confidence Intervals (CI)
- Standard error, *se*
- Critical values $z^*$
- Confidence levels (95%, etc)
- Margins of error, *moe*


## Topic 17: Tests of Significance: Proportions

So you've got a hypothesis, and you've sampled a population to prove it. Let's put your hypothesis on trial!

**Example**: Bio-psychologists conjectured couples have a tendency to turn to the right while kissing. Their team surveyed 124 couples at airports, train stations, beaches, and parks in U.S. and abroad. They found 80 of 124 couples leaned right when kissing. $\widehat{p}$ ?

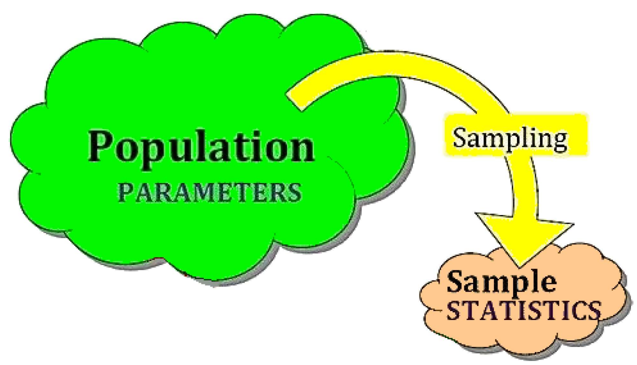
$\widehat{p} = \frac{80}{124} = 0.645.$

**Parameter** or **statistic**?

$\widehat{p} = 0.645$ is the **statistic** for this **sample**. But, this doesn't mean that $0.645$ is the proportion $\pi$ of ALL people who turn right.

And recall the original question was if more than ½ turn right.

Is $\widehat{p} = 0.645$ enough above $0.5$ to give evidence that $\pi > 0.5$?




## Siginificance Test

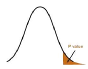
A **significance test** is a six-step procedure to test the likelihood of a hypothesis.

Putting our hypothesis on trial, where the sample & it's statistic are evidence.

Is the evidence strong enough to support the hypothesis?

We will go over...

## Six Steps of Significance Testing

1. Lay out the problem 
2. State hypotheses 
3. Check technical conditions 
4. Calculate test statistic ($z$-score) 
5. Calculate $p$-value 
6. Make a decision and write a conclusion 

---

## Step 1: Lay out Problem 

Define the:

- Parameter of interest (what symbol)?
- Variable to collect?
- Population to sample from?

**Example RQ**: Do more than ½ of couples turn right when kissing?

**Parameter**: Proportion of all couples who turn right when kissing. **Symbol**: $\pi$

**Variable to collect**: Does each couple turn right when kissing? **Population**: All couples.

---

## Step 2: State Hypotheses 


In a significance test, there are two hypotheses!

**Null Hypothesis** ($H_0$): current view of parameter (eg, assumed innocent null  )

**Alternative Hypothesis** ($H_a$): alternative view of parameter we're hoping to show (eg, guilty null  ).

For null ($H_0$), parameter $\pi$ is equal to some value $\pi_0$, denoted: $H_0 : \pi = \pi_0$.

Value $\pi_0$ could be any number between 0 and 1.

**From Example**: RQ - Do more than ½ of people turn right when kissing?

Null ($H_0$): ½ of people turn right: $H_0 : \pi = 0.5$.

Alternative ($H_a$): asserts $\pi$ is either:

- Less than $\pi_0$,
- More than $\pi_0$,
- Or not equal to $\pi_0$.

Alternative hypothesis looks like one (and only one) of the following:

$H_a : \pi > \pi_0$, $H_a : \pi < \pi_0$, $H_a : \pi \neq \pi_0$.

**Back to Example** (**RQ**): Do more than ½ of people turn right when kissing? Hypotheses?

$H_0 : \pi = 0.5$ and $H_a : \pi > 0.5$.

**One Sided vs Two Sided Tests**

When significance tests use an alternative hypothesis of:

$H_a : \pi > \pi_0$ OR $H_a : \pi < \pi_0$, it's called a **one-sided test**.

If sig. test uses $H_a : \pi \neq \pi_0$, it's called a **two-sided test**.

## Step 3: Check Technical Conditions

These are conditions required by CLT.

- Was data collected in unbiased way?
- Is sample size large enough? CLT demands $n\pi \geq 10$ and $n(1 - \pi) \geq 10$.

But we don't know $\pi$ !!


However, we have hypothesized value $\pi_0$ from $H_0$. So, we use that instead (note we don't use $\widehat{p}$).

**Back to Example**: Sample Size Check.

Researchers watched 124 couples.

Null: $H_0 : \pi = 0.5$.

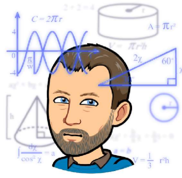
So: $n\pi_0 = 124(0.5) = 62 \geq 10$ and ✓

$n(1 - \pi_0) = 124(1 - 0.5) = 62 \geq 10$. ✓

Sample was also random, so technical conditions hold. ✓

---

## Step 4: Calculate Test Statistic ($z$)



Test statistic measures how far observed statistic $\widehat{p}$ is from hypothesized parameter value $\pi_0$.

If $\widehat{p}$ is "**far**" from $\pi_0$, then it's **likely** $H_0$ is **false**.

If $\widehat{p}$ is "**near**" to $\pi_0$ then it's **possible** $H_0$ is **true**.

**Test Statistic Derivation**

What's "far" or "near" is context dependent. However, we've seen how to create a consistent measure of distance by measuring it in terms of SDs. That is, the $z$-score.

We have a formula for SD w/a sampling distr: $\sqrt{\frac{\pi(1-\pi)}{n}}$.

We don't know $\pi$, so we'll use hypothesized $\pi_0$.

Note $\widehat{p} - \pi_0$ is how far statistic $\widehat{p}$ is from $\pi_0$.

So, test statistic $z = \frac{\widehat{p}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ is how far $\widehat{p}$ is from $\pi_0$ in terms of SDs.

The larger $z$ is, the less likely $\pi_0$ is to be correct.

**Example**: We have $n = 124$, $\widehat{p} = 0.645$, and $H_0 : \pi = 0.5$.

What is the SD?

$$s = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{124}} = 0.04490.$$

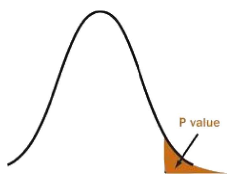
What is the $z$-score (test statistic)?

$$z = \frac{\widehat{p}-\pi_0}{s} = \frac{0.645-0.5}{0.04490} = 3.229.$$

So our $\widehat{p}$ of $0.645$ is $3.23$ SDs away from our hypothesized $\pi_0$ of $0.5$. Is this enough?

---

## Step 5: Find the $p$-value




Instead of just looking at distance, we can ask:
If $H_0$ is correct, what's the probability $p$ of observing this $\widehat{p}$ (or something more surprising)?

This probability $p$ is called the $p$-**value**.

**Back to Right-Kissing Example**: Recall $\pi = 0.5$ and $n = 124$.

What's probability of observing sample proportion $\widehat{p}$ of $0.645$ or higher?

"Or higher" because our alternative hypothesis is $H_a : \pi > 0.5$.

We already know $z$-score for $0.645$ is $3.23$. So, probability $p$ of $3.23$ or higher?

$z$-score calculator for $P(Z > z)$ is $0.0006190$.

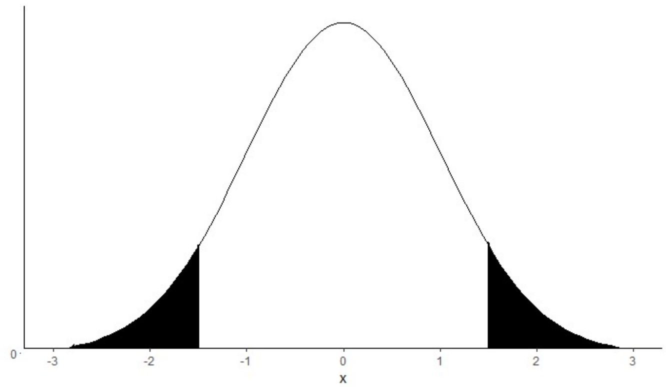
Probability of observing $\widehat{p} = 0.645$ or higher, assuming $\pi = 0.5$ and $n = 124$, is $0.00062$. ($0.062\%$)

---

## $p$-value Derivation

The probability $p$ that we calculate depends on alternative hypothesis $H_a$.

- If $H_a : \pi < \pi_0$, then we want probability of observing $\widehat{p}$ or something lower.
- If $H_a : \pi > \pi_0$, then we want probability of observing $\widehat{p}$ or something higher.
- If $H_a : \pi \neq \pi_0$, then we want probability of observing something at least as far from $\pi_0$ as $\widehat{p}$ is, but in either direction (more extreme than $\widehat{p}$).




2-sided $p$-value

---

## Step 6: Make Decision & Write Conclusion 

We first need to make a **decision**: Have we evidence that alternative $H_a$ is true (eg, guilty null  )?

Or do we not have enough evidence (eg, innocent null  ).

❗ We never can conclude null ($H_0$) is true based on our data.

This is because $H_0$ is an equality statement, and statisticians can't prove things are equal.

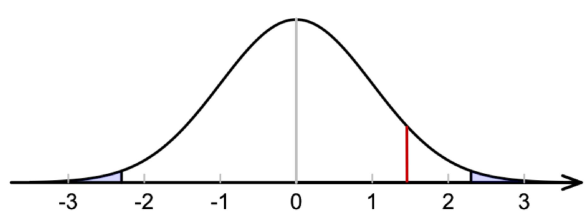
Then, **write a conclusion in context**, using no statistical symbols, about your decision.

**Decision**

Our $p$-value and test statistic ($z$) describe the strength of our evidence.

- Large test statistic $z$ & small $p$-value indicate $\hat{p}$ is far away from $\pi_0$.

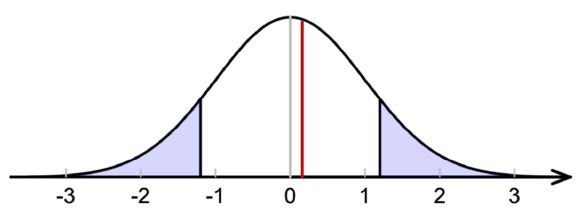
  Stat $\hat{p}$ is unlikely to be observed if $H_0$ is true.
  So, it's unlikely $H_0$ is true, and we **reject the null hypothesis**.




Small $p$ (purple), large $z$ (black line), $H_0$ is grey line, $p$ is red line

- Small $z$ & large $p$-value indicate $\hat{p}$ is relatively close to $\pi_0$.

  Stat $\hat{p}$ is likely to be observed if $H_0$ is true. So $\hat{p}$ is consistent w/$H_0$.
  Thus $H_0$ may be true, so we fail to **reject the null hypothesis**.




Large $p$, small $z$

- Small $p$-value, reject $H_0$ and conclude alternative $H_a$.
- Large $p$-value, fail to reject $H_0$ and fail to conclude $H_a$. (but **don't** conclude $H_0$ is true)

But how small is small enough?

**Significance Level**

*Before* beginning the experiment, we need to decide how small the $p$-value would need to be for us to reject $H_0$.

This $p$-value cut-off is called the **significance level** $\alpha$.

In most cases, $\alpha = 0.05$.

If $p$-value $< \alpha$, reject null $H_0$ and conclude alternative $H_a$.
If $p$-value $> \alpha$, fail to reject $H_0$ and fail to conclude $H_a$.

Small $p$, then REJECT!

---

**Example**: We observed $\widehat{p} = 0.645$ out of 124 couples while testing the hypotheses

$$H_0 : \pi = 0.5, \qquad H_a : \pi > 0.5.$$

We found test statistic of $z = 3.23$ and $p$-value of $0.00062$.

If significance level is $\alpha = 0.05$ (the default), what decision should we make?

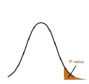
Since $p = 0.00062$ is less than $\alpha = 0.05$, we **reject the null hypothesis** $H_0$.
We conclude the alternative hypothesis $H_a$, that $\pi$ is more than $\pi_0 = 0.5$.

**Conclusion (in context)**: more than half of couples turn right when kissing. (use no stats symbols here)

---

In short....

## Six Steps of Significance Testing

1. Lay out problem 
2. State hypotheses 
3. Check technical conditions 
4. Calculate test statistic ($z$-score) 
5. Calculate $p$-value 
6. Make a decision and write a conclusion 

## Activities: 17-2

---

### What did we learn?

- Six steps of significance test